Open Government Portals Assessment: A Transparency for Accountability Perspective

Rui Pedro Lourenço¹,²

¹ INESC Coimbra, Portugal
² Faculty of Economics, University of Coimbra, Portugal
{ruiloure@fe.uc.pt}

Abstract. Dataset portals such as Data.gov and Data.uk.gov have become flagship initiatives of open government and open data strategies. These portals aim to fulfill the open government objectives of promoting re-use of public sector information to develop new products and services, and increasing transparency for public officials’ accountability. This work focus on the latter and its aim is to propose a set of requirements as part of a framework to assess whether dataset portals are indeed contributing to a higher degree of transparency focusing on accountability. Previous studies on internet-based transparency (including Internet Financial Reporting – IFR) were analyzed, from which several requirements were derived concerning the data types sought after, the public entities covered, the information seeking strategies adopted and the desired qualitative characteristics of data. The rationale behind our proposal is that dataset portals developed under the open government principles should, at least, be able to fulfill the informational and operational requirements identified in the ‘traditional’ transparency assessment literature.

Keywords: Dataset portals, transparency, accountability, assessment.

1 Introduction

The idea of open government is not new: Linders and Wilson [22] trace the origins of open government back to the Swedish Freedom of the Press Act in 1766. But perhaps the most notorious and visible recent effort to promote the principles of open government can be attributed to the Open Government Directive (OGD) [25], issued by the Obama Administration in 2009. Open government strategies such as the OGD include as one objective the promotion of governmental transparency which involves the “publication of government data intended to support accountability and reuse for social or economic value” [22]. In this work we are particularly interested in transparency for public accountability, that is, the disclosure of data which provides citizens and other public stakeholders “with the information needed for judging the propriety and effectiveness of the conduct of the government” [3].

Dataset portals (described as “services supporting the location of public sector information (PSI)” [33]) such as Data.gov and Data.gov.uk quickly became flagship initiatives of open government, but they are not isolated initiatives. Data.gov, for
instance, provides an extensive list of other open data sites, as part of open government initiatives, including 39 concerning US states, 39 related to US cities and counties, 40 maintained by other countries and 140 from international regions\(^1\). Huijboom and Broek [19] have also identified and analyzed open data strategies from five different countries (Australia, Denmark, Spain, the United Kingdom and the United States) which include the financing of open data portals.

Despite this widespread movement towards the promotion of governmental open data (dataset) portals, two questions arise: Are these portals effectively making data available for accountability (transparency) purposes? How can that be assessed? The aim of this paper is to identify and propose a set of requirements as part of an assessment framework that may be used to answer these questions.

The lack of evaluation guidelines regarding the implementation of open government principles was emphasized by Darbishire [10] who recognized that “limited examples of monitoring of proactive disclosure by official or oversight bodies” were found and recommended that monitoring of proactive disclosure initiatives should be put in place [10]. Even the OGD did not provide an assessment framework on how to evaluate US agencies plans and, more importantly, on how to evaluate their implementation and results [2]. More recently, Huijboom and Broek [19] noted that, from the five countries analyzed in their work, only the UK and the US have evaluated their open data policies but none of them have assessed their economic and social impacts. Harrison et al. [17] also acknowledge the lack of frameworks and procedures to assess open government.

Regarding the actual data being disclosed, McDermott [23] recognized that, in what concerns Data.gov, “no one has done an overall assessment of the data sets”. According to Harrison et al. [17], even if the data disclosed is “both usable and of high quality”, simply making data available does not necessarily mean that government is being more transparent and accountability of public agents is facilitated. As a consequence, these authors developed and proposed a “public value assessment tool” intended to guide assessment exercises and record the analysis product. In a similar effort, Reggi and Ricci [29] analyzed 434 EU Cohesion policy and Structural Funds related datasets, and classified them into three clusters according to their characteristics concerning the policy principles of "stewardship" and "usefulness" [11].

Another strand of assessment research dedicated to technical aspects of data disclosure, that is, the way data is being disclosed within dataset portals. Such research is usually focused on the respect for principles of Open Government Data [24] and Linked Open Government Data [33]. In this sense, Kalampokis et al. [20] recognize two broad technological approaches to classify Open Government Data and distinguish between “downloadable files” and “linked data”.

Despite all research efforts described, no dataset portals assessment framework were found, specifically in what concerns the ability of those portals to promote transparency and allow for accountability of public agents. The goal of this paper is therefore to provide such a framework in the form of requirements derived from the analysis of transparency research literature. Even before the emergence of open

government related dataset portals, such literature used assessment models and procedures that relied on individual web site analysis to evaluate the transparency degree of public entities. The assessment models used, and their underlying informational and procedural requirements, should apply to the new context of ‘dataset portals’ by defining a minimum set of requirements dataset portals must fulfill to respond to the needs of citizens and other stakeholders when conducting accountability processes. Therefore, these requirements form the basis of our assessment framework proposal.

The remainder of this paper is structured in the following way. Section 2 briefly describes the transparency assessment literature considered. Section 3 analyses those papers according to 4 relevant dimensions which will be considered in section 4 to propose a set of requirements that compose the assessment framework. Some final remarks are presented in section 5 (conclusions).

2 Transparency Assessment Literature

The proposed framework assumes that dataset portals, in order to promote government transparency and accountability, should (at least) fulfill the requirements identified in the transparency assessment literature. Table 1 presents the list of research papers considered (the more recent works are presented in the end), the majority of which address internet-based transparency, with one exception [7] (which considers paper-based disclosure). The list does not result from a formal literature review process, but rather from an ongoing research effort.

3 Analysis

To analyze these papers two important dimensions were considered: the type of public entities studied and the information types sought-after. The sequence by which these studies define those two dimensions characterize different information seeking scenarios. Once these two dimensions are established, studies also consider in their assessment technical and qualitative aspects of the data being disclosed. Only the latter will be considered in our analysis.

3.1 Entities Covered

Several types of entities were addressed by these studies, including different types of local, regional and central authorities, and universities. Clearly, local governments are the most chosen research target. Some studies address a single type of entities while others consider several (although similar) entity types. In some cases the authors perform international comparisons, although most research concentrates on entities from a single country. The number of entities considered in each study varies. Some studies consider all entities of a particular type (for entity type characterization purposes), while others select a sample of entities of a particular type based on some characteristic criteria (e.g. [8]) or using a random selection process (e.g. [35]).
Table 1. Transparency assessment literature

<table>
<thead>
<tr>
<th>Research paper</th>
<th>Entities</th>
<th>Information type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coy and Dixon [7]</td>
<td>New Zealand universities</td>
<td>Annual reports</td>
</tr>
<tr>
<td>Groff and Pitman [15]</td>
<td>100 Largest U.S. municipalities</td>
<td>Financial reporting (budgets, CAFRs, popular reporting)</td>
</tr>
<tr>
<td>Caba Pérez et al. [4]</td>
<td>Central administrations of the E.U. countries</td>
<td>Budgetary and financial information; Management indicators</td>
</tr>
<tr>
<td>Laswad et al. [21]</td>
<td>30 Local authorities in New Zealand (regional, city, district councils)</td>
<td>Financial highlights; Annual reports; Annual plans</td>
</tr>
<tr>
<td>Rodríguez Bolívar et al.</td>
<td>Anglo-Saxon, South American and Continental European central governments (Departments of Finance)</td>
<td>Budgetary and financial information; Management indicators (performance)</td>
</tr>
<tr>
<td>Rodríguez Bolívar et al.</td>
<td>Spanish regional governments</td>
<td>Economic and financial information; Budgetary information; Performance indicators</td>
</tr>
<tr>
<td>Pina et al. [26]</td>
<td>319 EU regional and local governments (76 cities and 242 regions or counties of the first 15 EU countries)</td>
<td>Transparency (Ownership, Contacts, Organizational or operational information, Issues, Citizen consequences, Security and Privacy)</td>
</tr>
<tr>
<td>Styles and Tennyson [36]</td>
<td>300 U.S. Municipalities</td>
<td>Comprehensive Annual Financial Reports</td>
</tr>
<tr>
<td>Caba Pérez et al. [5]</td>
<td>Spanish municipalities</td>
<td>Budgetary information and cash-flow; Financial position; Non-financial information (Indicators related to public service management: efficiency, effectiveness, and economy)</td>
</tr>
<tr>
<td>De Kruijf [12]</td>
<td>Municipalities above 40,000 inhabitants in North Rhine Westphalia and the Netherlands</td>
<td>Full budget documents; Annual reports</td>
</tr>
<tr>
<td>Gandia and Archidona [13]</td>
<td>Spanish municipalities above 50,000 inhabitants</td>
<td>General information; Budgetary information; Financial Information</td>
</tr>
<tr>
<td>Research paper</td>
<td>Entities</td>
<td>Information type</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>Serrano-Cinca et al. [32]</td>
<td>Spanish provincial capitals and all town halls with over 70,000 inhabitants</td>
<td>Financial information (budget or annual accounts)</td>
</tr>
<tr>
<td>Pina et al. [27, 28]</td>
<td>75 European local governments (the 4 biggest cities plus the capitals of the first 15 E.U. countries)</td>
<td>Economic and financial information; Performance, social and environmental information; Transparency (organization chart, mission statement/activities, …)</td>
</tr>
<tr>
<td>Cárcaba García and García-García [6]</td>
<td>334 Spanish municipalities above 20,000 inhabitants</td>
<td>Informative content I (Budget, Balance sheet, Income statement, …); Informative content II (Explanations, Benchmarking references, …)</td>
</tr>
<tr>
<td>Yu [37]</td>
<td>Local government in China (all 63 provinces and cities excluding Hong Kong, Macao and Taiwan)</td>
<td>Idem [36]</td>
</tr>
<tr>
<td>Guillamón et al. [16]</td>
<td>100 Largest Spanish municipalities</td>
<td>Items from the Financial Transparency Index created by Transparency International Spain</td>
</tr>
<tr>
<td>Reggi and Ricci [29]</td>
<td>Regional operational programmes and regional managing authorities from EU member states</td>
<td>Data concerning projects and beneficiaries of the European Regional Development Fund (ERDF) and the European Social Fund (ESF)</td>
</tr>
<tr>
<td>Armstrong [1]</td>
<td>Florida's 67 counties and the 67 corresponding school districts</td>
<td>12 Public records in accordance to Florida State Statutes, including government contracts, budgets or financial information, and meeting agendas.</td>
</tr>
<tr>
<td>Grimmelikhuijsen and Welch [14]</td>
<td>80 Dutch municipalities</td>
<td>Air quality data concerning the transparency of decision making, policy information and policy outcome</td>
</tr>
<tr>
<td>Sol [34]</td>
<td>Spanish municipalities</td>
<td>Idem [16]</td>
</tr>
<tr>
<td>Cucciniello et al. [8]</td>
<td>Pilot (purposive) sample of 113 Italian cities (Italian provincial capitals)</td>
<td>Information concerning four dimensions (institutional, political, financial, and service delivery), each of which concerning three stages (static, dynamic, feedback).</td>
</tr>
</tbody>
</table>
3.2 Information Types Covered

Noticeably, the most sought-after data is related to financial, budgetary and management information (including performance indicators). More detailed data concerning public spending (still with a financial impact), such as government contracts [1] and public funds management programs [29], is also considered.

Cucciniello et al. [8] add to this financial dimension three other types of information associated with a political, service delivery and institutional dimension. The latter is based on the required items suggested in the Web Site Attribute Evaluation System (WAES). This was developed by the Cyberspace Policy Research Group ([9] apud [27]) and was originally used for evaluating US federal websites [27]. This institutional dimension is adapted and used on several works ([8, 26-28]) and it refers to items such as ownership (of web site), contacts, or activities (static information [8]). Governance information (such as strategy, planning and capability building) is also considered in the studies [35]. Finally, Grimmelikhuijsen and Welch [14] used the process and event transparency concepts proposed by Heald [18] and search for data concerning decision-making processes, policy content, and policy outcomes.

From the different main types of information identified in our survey it is possible to conclude that the nature and type of the information sought-after in transparency studies is not always ‘compatible’ with the dataset format: sometimes researchers look for mainly textual reports with little supporting (numerical) data. Also, some data is clearly produced in continuum (such as contract related data) while other is produced at a single point in time (such as budgets).

3.3 Information Seeking Strategies

In general, research papers address a particular type (universe) of entities and aim at assessing the degree of web-based transparency of all or part of entities belonging to that universe by seeking on their individual web sites for relevant information items (and the way they are disclosed). The work by Reggi and Ricci [19] is somehow distinctive in the sense that it begins by identifying the type of information sought after (European Regional Development Fund and the European Social Fund), and then identifies the entities that have the responsibility, at national level, to disclose it.

Although the common goal of these studies is to assess transparency, some aim to perform a descriptive analysis of the information disclosed (which entities do not report the expected items, etc) [29, 35], others aim to quantify the level of transparency and build a disclosure index (e.g. [31]), while still others identify factors (disclosure incentives) that account for the levels of measured transparency (e.g. [6]). Despite some differences, the authors usually adopt the following general procedure:

1. Select and define the type of entities to be assessed;
2. Identify a subset or the entire universe of entities of that particular type. In the former case, selection may be based on a specific characteristic (such the number of inhabitants of a municipality [13]) or entities may be randomly selected (e.g. [35]);
3. Determine and characterize the information item(s) expected to be disclosed by each of the entities considered;
4. Define other analysis dimensions of data availability, including desired qualitative characteristics, and technical characteristics of data disclosure and/or web site;
5. Define the global assessment model which considers all analysis dimensions, including determining weights for the different criteria and developing a scoring methodology (if applicable);
6. Visit each entity web site, and collect the data required by the assessment model to characterize or evaluate the entity.
7. Describe the results or compute the disclosure indexes according to that model.

As mentioned before, a slightly different information seeking scenario would start by identifying the required information types (step 3) and only then determine which types of entities should be assessed (steps 1 and 2): typically those responsible for the information considered. Dataset portals assessment frameworks must take into consideration these information seeking scenarios and other possible variations.

3.4 Qualitative Requirements of Data

The assessment exercises described in the literature usually consider in their models qualitative characteristics of the data available and technical characteristics of the web sites. The required qualitative characteristics of data are commonly selected based on international best practices (e.g. Caba Pérez et al. [4]) and most commonly include the requirements for completeness, timeliness, understandability or clarity, comparability, relevance and reliability [4, 30]. Some of these qualitative requirements are also considered as part of the eight principles of Open Government Data [24].

4 Assessment Framework Requirements

Based on the analysis presented in previous section, it is now possible to propose a set of requirements that form an open government dataset portals assessment framework. These requirements will be associated with each of the three main topics addressed in the previous section (entities covered, information types, and information seeking strategies). Whenever relevant, references will be made to desired qualitative characteristics of data although strictly technical aspects of data provision, such as data formats or linked data principles will not be considered.

4.1 Entities Covered

According to the literature analyzed, dataset portals, in order to fulfill transparency informational needs, should address the whole range of entity types possible in each country institutional arrangement. Portals curated by international organizations, might even be expected to cover entities from different countries. In this respect, users should be able to assess the completeness of the data provided.

R1: Portals should present a master list of all entity types covered (regardless of the entities that actually have provided datasets).
This means that dataset portals should cover the entire set of entities of a particular type. Again, in this context, users should be able to assess the completeness of the data provided. Some portals do indeed present a list on entities which have actually disclose datasets, but that information is not sufficient to identify which entities should provide the same information but are not doing so.

R2: Portals should present a master list of entities belonging to each one of the types covered (regardless of the entities that actually have provided datasets).

Search mechanisms alone do not meet the requirements identified in the transparency literature. A browsing mechanism, based on the master list of entities, offers a more reliable way to determine which entities, from each type, do not provide data.

R3: Portals should present a list of disclosed datasets associated with each entity.

4.2 Information Types Covered

To fulfill transparency informational needs recognized in the literature, portals should provide all datasets related to, at least, the whole range of information types identified in the previous section. These datasets should be clearly identified and associated with a specific type of transparency item. It should also be possible to easily determine which information items are not being covered by the datasets or are not being provided by a certain entity (for instance).

R4: Portals should provide a master list of accountability-oriented information items, and related datasets, which should be available (regardless of the actual provided datasets).

Some portals allow users to characterize their datasets using tags such as “transparency” or “accountability” (e.g., Data.gov) which are not consistently used and do not provide a complete view of all transparency/accountability related information provided. This master list should clearly distinguish the different types of transparency datasets (accountability or re-use).

R5: Portals should associate each dataset disclosed with each information type defined.

The literature also showed the importance to access institutional (entity characterization) information needed to allow for characteristic-based selection of entities.

R6: Portals should provide, for each entity, a minimum of institutional (characterization) information regardless of the other datasets provided by the entity.

Transparency literature also showed the importance of ‘non-dataset compatible’ information (such as textual reports) to accountability. In the context of a broader open government policy, portals should be adapted or complemented to cope with such type of information, despite the limitations of their current specific dataset nature.
R7: Portals should be able to accommodate ‘non-dataset compatible’ transparency-related information items (such as reports and other documents).

The literature also revealed the importance of clearly knowing the date and time associated with each dataset, and to follow the evolution of indicators (for instance). Sometimes, entities provide several datasets concerning the same information item but related with different time periods (years, for instance).

R8: Portals should explicitly indicate, for each information type, which update periodicity is expected for individual datasets.

Also, without and explicit dataset time organization structure, it becomes very difficult to identify missing datasets from a particular entity at a given time, and to perform longitudinal analysis concerning a particular entity or information item.

R9: Portals should explicitly associate each dataset with a specific time/period tag.

Whenever possible and relevant, entities should provide the data as it is created and not only on a periodic (yearly) base, thus allowing for real-time transparency [18]. The timeliness of the data is one important qualitative characteristic valued in the literature.

R10: Portals should be prepared to allow datasets to be updated in real-time, thus reflecting the circumstances at each moment.

The assessment exercises reported in the literature rely on the comparability of data provided by different entities. This implies that for each information type, a common set of characterizing elements must exist to allow for such comparability.

R11: Portals should define the minimum set elements that must be disclosed by each entity type, for each information type.

This set of characterizing elements should also take into consideration the principles of completeness and relevance, two qualitative characteristics of data identified in the literature.

4.3 Information Seeking Strategies

Different possible information seeking strategies were identified from the literature. Portals, as key components of open government initiatives, should support citizens to pursue accountability regardless of the way information is sought-after.

R12: Portals should adopt and make visible an overall organization structure, according to the requirements previously presented, and provide related browsing and selection tools.
This organization structure and corresponding tools should facilitate the identification of non-reporting entities, missing datasets by some entities and longitudinal studies (for instance).

5 Conclusions

There is a significant body of scientific literature dedicated to the assessment of web-based public entities transparency. Most of this literature considers the data provided in each entity individual web site as a proxy for its transparency. But the emergence of dataset portals, in the context of open government programmes, has an impact on the way transparency is ‘traditionally’ assessed: from now on it is necessary to consider that at least part of the data expected to be disclosed in an entity web site might be available through ‘collective’ dataset portals. This means that the procedures adopted by this assessment literature might no longer be totally adequate.

Also, dataset portals were created to meet the objectives of open government strategies, including promoting transparency and therefore facilitating the accountability of public agents. However, no evidence was found in the open government literature that the way data is provided by these portals, and the data itself, might in fact contribute to meet such objective. Research has been mainly focusing on open government strategies formulation, implementation and impact assessment, and technical aspects of data availability. One matter seems clear: disclosing huge amounts of datasets does not necessarily equate to more transparency and does not necessarily facilitate accountability.

Despite the shortcomings of the assessment methodologies identified in the literature to deal with the new portal reality, they may still be considered as proxies for the way information is sought-after in the context of accountability processes. Therefore, this work analyzed the ‘traditional’ transparency assessment literature and identified a set of requirements that dataset portals need to fulfill in order to contribute to the transparency of public entities and allow for the accountability of public officials. Such requirements concern the type of entities covered by dataset portals, the type of information types provided, the information seeking strategies supported and some qualitative aspects regarding the data provided.

Further research is needed to complement the proposed set of requirements by considering, for instance, qualitative aspects of data delivery and the informational needs of ‘real’ citizens in what concerns public accountability. The next step is to fully develop an assessment model and procedure based on the requirements identified which could be used to evaluate dataset portals and other elements of open government strategies. The ultimate goal is to give governments the information they need to improve their open government strategies and therefore contribute to increase transparency and allow for better accountability.

Acknowledgements. This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) under project grant PEst-C/EEI/UI0308/2011.
References
